DEVELOPMENT OF BREAST CANCER DETECTION SYSTEM USING DIGITAL IMAGE.

NORLAILAH BINTI LANISA

A report submitted in partial fulfilment of the
requirements for the award of the degree of
Bachelor of Electrical Engineering (Electronic)

Faculty of Electrical and Electronic
Universiti Malaysia Pahang

JUNE 2012

ABSTRACT

Breast cancer is one of the most common cancer affecting women around the world. Mammography is the most known and effective method to detect early signs of breast cancer. However, due to some weaknesses in mammography such as painful procedures and radiation, researches introduce another diagnosis method which is by analysing thermal image. The purpose of this project is to design a system to detect the signs shown in mammogram and thermal image using Image Processing Technique applied to MATLAB. Image processing techniques can be divided into several elements. The elements are image acquisition, image pre-processing, image processing, feature extraction, object classification and classification decision. Both type of images then analyse step by step according to the elements. Mammogram images are analyse using morphology technique before features extraction which then lead to the classification of the image into 3 classes (‘Normal Fatty breast’, ‘Abnormal Fatty breast’ and ‘Glandular breast). For thermal image, the distribution of heat around the breast will be the features extracted and analysed. The different range of heat in the image will be used to specify the possible area of cancer. This project also includes the construction of Graphical User Interface (GUI) so that the system is more users friendly.

ABSTRAK

Barah payudara adalah sejenis barah yang paling biasa dialami oleh ramai wanita di seluruh dunia. Teknik Mammografi adalah teknik yang paling kerap dilakukan dan dikatakan paling efektif dalam mengesan tanda-tanda awal barah payudara. Walaubagaimanapun, prosedur yang menyakitkan dan radiasi dikenalpasti sebagai kelemahan teknik mammografi sehingga menyebabkan para penyelidik memperkenalkan teknik diagnosis mengunakan imej suhu. Projek ini dilakukan dengan tujuan untuk mereka satu sistem mengunakan teknik pemprosesan imej pada MATLAB yang mampu mengesan tanda-tanda barah berdasarkan kepada imej mammografi dan imej suhu payudara. Teknik pemprosesan imej boleh dibahagikan kepada beberapa komponen. Komponen yang terlibat adalah pemerolehan imej, imej pra-pemprosesan, pemprosesan imej, pengekstrakan maklumat, pengelasan objek dan keputusan klasifikasi. Kedua-dua jenis imej kemudian dianalisis mengikut komponen. Imej mammogram dianalisis menggunakan teknik morfologi sebelum maklumat imej diekstrak untuk mengklasifikasi imej kepada 3 kelas yang berbeza ('Payudara Lemak Biasa', 'Payudara Lemak Abnormal' dan 'Payudara Kelenjar'). Bagi imej suhu, pengedaran haba di sekeliling payudara adalah ciri-ciri yang akan diekstrak dan dianalisis. Perbezaan haba dalam imej akan digunakan untuk menentukan kawasan yang berkemungkinan barah. Projek ini juga melibatkan pembinaan Graphical User Interface (GUI) untuk mewujudkan satu sistem mesra pengguna.

TABLE OF CONTENT

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

Breast can be classified into 2 types due to its density which are ‘Fatty breast’ and ‘Glandular breast’. When the amount of fat tissues exceed the amount of fibro-glandular tissues, the breast can be classified as ‘Fatty breast’ and when the amount of fibro-glandular tissues exceed the amount of fatty tissues, the breast can be classified as ‘Glandular breast’. Breast cancer occurs when breast tissues grow, change and multiply rapidly without control which may form lump or mass of extra tissues as shown in Figure 1.1. These masses are called tumor and can be either cancerous (malignant) or non-cancerous (benign) [1].

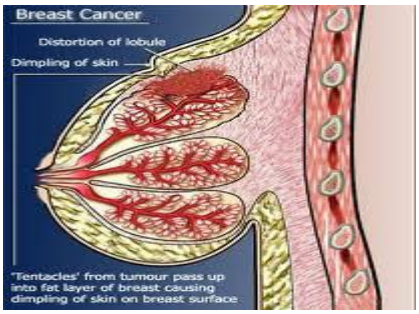

Figure 1.1 Breast cancer

Breast cancer is one of the most common cancers affecting woman and the most common source of death among middle aged women. Based on the World Health Statistics 2011 by Global Health Observatory (GHO), the mortality among female population all over the world cause by malignant neoplasma is about 11.81% and breast cancer are the highest with 15.80% compare to other types of cancer [2]. Successful treatment of breast cancer depends on early detection. Currently, two imaging method uses to detect masses are mammography and thermography.

1.1 Problem Statement

Mammography is a specific type of x-ray imaging that focusing on breast imaging. This process uses system with low dose x-ray, high contrast and high-resolution film [3]. However the accuracy in detecting the cancer based on mammogram image with bare eye by qualified personnel (radiologist) will be affected by poor mammographic image quality and fatigue radiologist. Radiologist misdiagnose 10-30% of the malignant cases due to the difficulty to maintain required attention level when reading large number of screening mammograms since most are free of cancerous features [1]. However, due to the invasive process of mammography, another method emerged as potential method to detect the early sign of breast cancer. The method is thermography.

Thermography is an imaging technique which shows temperature patterns at the surface of the breast. The heat patterns indicate the metabolic activity and vascular circulation in breast tissue which will be high at the surrounding area of cancerous tissue [4].

Due to the need of overcoming the problem that cause high rate of false positive and false negative detection, a Computer Assisted Detection (CAD) system is develop to provide assistant for clinician to identify cancerous tissue in mammogram and thermal image. The system will be design based on image processing technique on MATLAB platform.

1.2 Objectives

This project proposes to develop a system for breast cancer using image processing technique.

1.2.1 Sub-objective

The sub-objective for this project is as follow.

1. To classify type of breast and detect abnormalities of breast using mammogram image.
2. To study thermal image processing feasible to detect breast cancer.

1.3 Project Scope

The 150 breast mammogram images for this study are obtained from trusted online database (MIAS database) [11]. Breast thermal image with abnormalities obtain from 6 case studies by Pacific Chiropractic and Research Centre Infrared Imaging in California [4]. Both type of image analyse using Matlab software.

CHAPTER 2

LITERATURE REVIEW

Several researches have been done to develop CAD system to detect breast cancer. The references for this paper are taken from journal, books and conferences regarding the mammogram image and thermal image.

2.1 Image Processing Technique

Alasdair McAndrew (2004) entitlement, image processing is used to change the nature of an image to improve and enhance the image for human interpretation. Image processing also used to render image for machine perception. In his module, he explains on how to use matrix capabilities of MATLAB to investigate images and its properties. The image processing operations are explained in term of chapters. Image display chapter explain on how to use *imshow* function to display image and how spatial resolution and quantization affect the display and appearance of the image. Another chapter explain on point processing and the sub-chapter are arithmetic operation, histogram and thresholding. This sub-chapter discuss on how to modify image (enhance and blurring image) using MATLAB function that show how each operation work. For example, *imadjust* function which indicate histogram stretching is use to enhance image. The next chapter teach about spatial filtering. This chapter is also explaining on how to enhance and blurring image but using different operation. The operation discuss in this chapter is by filtering image using frequencies (low and high pass filter), Gaussian filter and non-linear filter. Types of noise, cleaning salt and pepper noise and cleaning Gaussian noise are explain in the

next chapter which is noise chapter. Noise is degradation in the image cause by image disturbance during transferring and during image acquisition. Cleaning noise is important to restore image to its original state and to analyse the image. The type of noise discuss in this module are salt and pepper noise and Gaussian noise. The image is filtered using *fspecial* function to clean up the noise. The next chapter explain one of the most useful information in an image which is edge. The uses of finding the edge are to measure size of the object, to isolate object from background, to recognise and classify object in the image. There are numbers of edge detection method discuss in this module such as Robert edge detection, Sobel edge detection and Prewitt edge detection. The module also discuss on topic morphology. Morphology is an operation in image processing to analyse shape in image. Morphology consist of many types of operations and some of it such as dilation, erosion, opening, closing, hit or miss transform, region filing and connected components are discuss in this module. Topic colour processing is discussed in the next chapter. In this chapter, the main topic discuss are what is colour mean in image processing, colour models, colour image in MATLAB, pseudo colouring and colour images processing . For example, to extract RGB component in RGB image, *imshow* function can be used. [17]

Marius Leordeanu et. al. (2011) entitlement, boundary detection is an important task when doing segmentation and recognition using vision system. Despite being one of the most important tasks to be done in image processing, there is no general formulation for boundary detection. This paper discusses on formulation and algorithm designed to detect different types of boundaries such as boundaries intensity, occlusion boundaries and specific boundaries for object. Based on their observation, boundary can be summarizing as a region that separates different image regions or a layer that coincide with boundaries in other layer. In this paper, author has designed two algorithms (Gb1 and GB2) and tested the algorithm to detect boundaries in static colour images, occlusion boundaries in video, occlusion boundaries in RGB-D video and boundaries from soft-segmentation. The result is the algorithm effectively and accurately detects boundaries of images use in the experiment. [18]

Hao Yuen Kueh et. al. entitlement, biological image contain a lot of patterns and objects which may convey information about biology mechanism. This tutorial

discusses the process to extract data from raw microscopic image using MATLAB. The advantages to extract and quantify objects and patterns using automated image analysis compare to manual methods of analysis is automated image analysis will provide unbiased approach to extract information from images and testing hypotheses. Automated system analysis also has advantages to facilitate the collections of large data collections for statistical analysis. The topics discuss in the first section of this tutorial are how to read, display, write and convert images. Besides that, the author also discuss on how MATLAB represent image and how to convert between different types of image. The second section discuss on contrast adjustment. As majority of biological image have low dynamic range and the features are difficult to be analysis, there is a need to enhance the appearance of the image by using different intensity transformation. This step may improve the performance of image segmentation algorithm and feature recognition. Next section discuss on spatial filtering technique. The filters explained are smoothing filters (average filter and Gaussian filter), edge detection filter (Prewitt filter and Sobel filter), Laplacian filter and median filter. Mathematical morphology which uses to extract features and components in images discussed on the next chapter. The operations are dilation, erosion, opening (erosion followed by dilation), closing (dilation followed by erosion), filling holes and clearing border objects. Image segmentation process to subdivide image into regions and images discuss in the next section. The quantitative information is processed and analysed using segmentation technique for extraction. The techniques of segmentation are edge detection and morphological watershed. This tutorial also discuss on analysis of dynamic and motion in biological images. The techniques use to visualize dynamical behaviour are kymographs (two-dimensional analog of times traces), difference images, maximum intensity projections, image cross-correlation and particle tracking. [19]

2.2 Mammogram Image Analysis

Jelene Bozek, Kresimir Delac and Mislav Grgic (2008) entitlement, mammography is the best method to detect early signs of breast cancer such as masses, calcifications, bilateral asymmetry and architectural distortion. However,

due to human limitation computer system have to take the major role in detecting abnormal tissue. The challenges that have been faced by the system are the wide range of abnormalities features and the indistinguishable from surrounding cell. Most of system developed involves algorithms which consist of two stages. The first stage is to detect suspicious lesion and second stage is to reduce the number of false positives. In BI-RADS system which discussed in this paper, the detected lesions are classified as masses, calcifications, architectural distortion and bilateral asymmetry. Masses are classified as benign or malignant based on density (fat containing, low density, isodense and high density), margins (circumscribed, microlobular, obscured, indistinct and spiculated) and shape (round, oval, lobular and irregular). Calcifications classified as benign, malignancy suspicious and malignancy highly suspicious based on the distribution of cluster, size, shape, and variability. For architectural distortion, the lesion classified as malignant when integrated with other lesion such as masses and classified as benign when scar and self-tissue damage due to trauma detected. Bilateral asymmetry analyse based on its texture, shape measurement, topology, brightness distribution, roughness, pattern assymetry and directionality. [3]

Ranjeet Singh Tomar et. al. (2009) entitlement that image processing techniques that been mention in their journal are more radiologist friendly. The system is designed using image processing technique on MATLAB. The techniques uses are edge detection and morphology. The detection process designed will start with detecting entire cell in the image, followed by filling gaps, dilating gaps, removing border, smoothing the objects, finding structures and lastly extracting large objects. For feature extraction to find the wanted area, 3 steps were suggested which were reduce uneven illumination, determine size distribution in Top-hat Image and calculate first derivatives. The result for feature extraction then plotted into graph to be analysing for classification. [6]

Hala Al-Shamlan and Ali El-Zaart (2010) entitlement, the features extraction in mammogram is an important key for early detection of breast cancer. In this study, they aim to determine the features extraction range. Before the range was determined, image pre-processing and image segmentation process applied to the images. Image pre-processing done to the image to suppress noise and improve the contrast of the image. Image segmentation is for detect the suspicious lesion. The

features used are based on three main categories; Geometric, Texture and Gradient features. For Geometric category, the features measured are area, perimeter and compactness [20]. Features in Texture category mostly obtain from image histogram. The features are mean, mean global area, mean local area, uniformity, standard deviations, smoothness, skewness, entropy, correlation and inverse. The last category is Gradient category. Features that classified under this category are Sobel-mean, Sobel-mean global area, Sobel-mean local area, Sobel-uniformity, Sobel-standard deviations, Sobel-smoothness, Sobel-skewness, Sobel-entropy, Sobel-correlation and Sobel-inverse. After applying up to 23 types of features extraction to 80 mammograms, they manage to obtain the range value for each feature extraction which may be used for further process in their breast cancer CAD system. [7]

2.3 CAD System Using Thermal Image

Monique Frize, Christophe Herry and Rober Roberge (2002) entitlement, the 3 technique in Head et al's methods [21]. The study shows the third method provided reliable result compare to the first and second method when applied to 9 patient's sample (6 with a diagnosis of normal and 3 with cancer). One of the analysis done is by increasing the threshold value in the methods and the result obtain are no false negatives or false positives on the sample. Therefore by looking at this preliminary result, they concluded future work should focus on improving third method to enhance thermogram diagnosis and decrease false negatives or false positives. [8]

V. Umadevi, S. V. Raghavan and Dr. Sandeep Jaipurkar (2010) entitlement, an interpretation system able to characterize thermography image as normal or required follow-up with clinician. The infrared cameras uses in this paper are Ti40FT from M/s Fluke Corp. and Varioscan-3021 ST from Jenoptik Laser. Software that integrated with the cameras to view images captured are SmartView for Ti40FT and IRBIS for Varioscan 3021-ST. The system discussed in this paper is Infrared Thermography Based Image Construction (ITBIC) system. This system consists of two main process, body boundary identification and highest temperature area extraction. This system will classify the image into normal case or follow-up case. When the system tested to 50 female volunteers, the system manages to characterize

and match the result to clinical finding. The system then interface with developed graphical user interface (GUI) to allow easier thermal image analysis by the radiologist or clinician. [9]

Pragati Kapoor and Dr. S. V. A. V. Prasad (2010) entitlement, image segmentation with automatic approach may improve the accuracy in earlier detection of breast cancer for thermography image. The methods outlined in this study are image segmentation and asymmetry analysis. One of the image segmentation process that use in this paper is edge detection which extract the boundaries of the breasts. The process also involving Hough transforms to extract the lower breast boundaries. Segmented classification done to classify each segmented pixel into a certain number of clusters. Lastly, diagnostic based on asymmetric analysing of the pixels in every cluster. The features uses in this paper for the diagnosis of image are skewness and kurtosis. [10]

CHAPTER 3

RESEARCH METHODOLOGY

This study will be using image processing main elements which are image acquisition, image pre-processing, image processing, feature extraction, object classification and classification decision as shown in Figure 3.1 for developing CAD system for mammograms and thermal images.

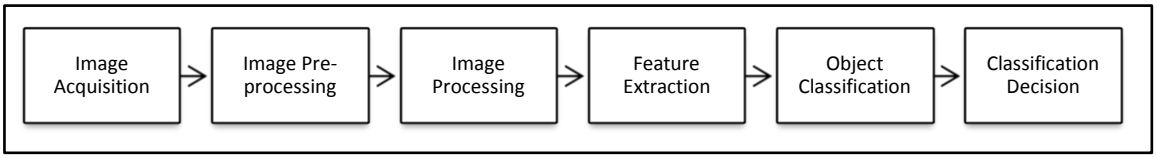

Figure 3.1 Image processing main elements

Image acquisition step involves the camera and its connection to the computer or processors. Computer or processors will receive the image in digital format. Image pre-processing step is a step to improve and enhance the image for processing step. Image processing step is a further step in analysing the image to obtain desired object. A lot of image processing techniques can be used in this step such as morphological processing, edge detection and compression. Feature extraction is where a set of desired features extracted from data pixels of the image which are good for classification. Object classification and classification decision are steps to make decision based on test and analysis done on the image [12]. The above methods is use to develop the following system;

1. System 1: Classification between ‘Fatty breast’ and ‘Glandular breast’.
2. System 2: Detection for abnormalities in type ‘Fatty breast’ images.
3. System 3: Detection of high temperature in thermal images.

3.1 Image Acquisition

3.1.1 System 1.

Mammogram image is obtained using mammogram machine. There are many types of mammogram machine and one of it is ‘YSX0906 Digital Mammography X-ray System’ manufactured by YSENMED from China as shown in Figure 3.2. The sizes of mammogram image that can be obtained using this machine are 18 x 24 cm or 24 x 32 cm. The C-arm of the machine can rotate between range +180° and -135°. The clinician is allowed to magnify the image with ratio from (1.4)/1 until (1.6)/1. The thickness of breast that can be compress from 0 mm until 268 mm with pressure range between 0 kg until 20 kg [13].

Figure 3.2 YSX0906 Digital Mammography X-ray Machine

In this study, The CAD system will be tested using 150 mammography images (65 ‘Fatty breast’ image and 85 ‘Glandular breast’ image). The digital mammography images acquired from online mammogram database (MAIS database). The image resolution of the image is 1024 x 1024 and in PGM (Portable Graymap) format. A sample of the database is shown in Figure 3.3.

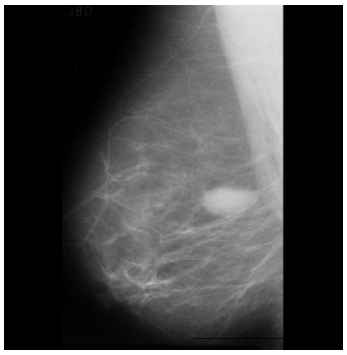

Figure 3.3 Mammogram image

3.1.2 System 2.

For detection for abnormalities, the system will test the 65 ‘Fatty breast’ mammography image that had been classified among the 150 mammogram images tested before. The digital mammography images acquired from online mammogram database (MAIS database). The image resolution of the image is 1024 x 1024 and in PGM (Portable Graymap) format.

3.1.3 System 3.

Thermal images of breast can be obtained using infrared camera. One type of camera that can be used is FLIR A615 manufactured by FLIR Systems, Inc. as shown in Figure 3.4. FLIR A615 is a perfect instrument for industries when the temperature changes over time is quiet fast. FLIR A615 also complies with standards like GigE Vision that allow this camera to interface using the Gigabit Ethernet communication protocol and fast image transfer using low cost standard cables even over long distances. This camera also complies with GenICam protocol which allows the camera to be use with third party software. Due to its compliance to standards, FLIR A615 is a Plug&Play device within 3rd parties Machine Vision softwares like NIs IMAQ Vision™ and the MVTecs Halcon™ software. By using this camera, image with resolution 640 x 480 pixels can be obtained [14].

Figure 3.4 FLIR A615 infrared camera.

For thermal image analysis, the system will test the 6 image at front position that taken from 6 case studies by Pacific Chiropractic and Research Centre Infrared Imaging website. The image resolution of the image is around 300 x 200 and in JPG (Joint Photographic Expert Group) format. A sample of thermal image is shown in Figure 3.5.1.

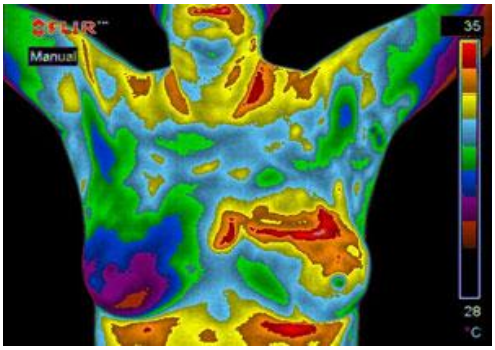

Figure 3.5 Thermal image of breast

3.2 Image Pre-processing and Image Processing

3.2.1 System 1.

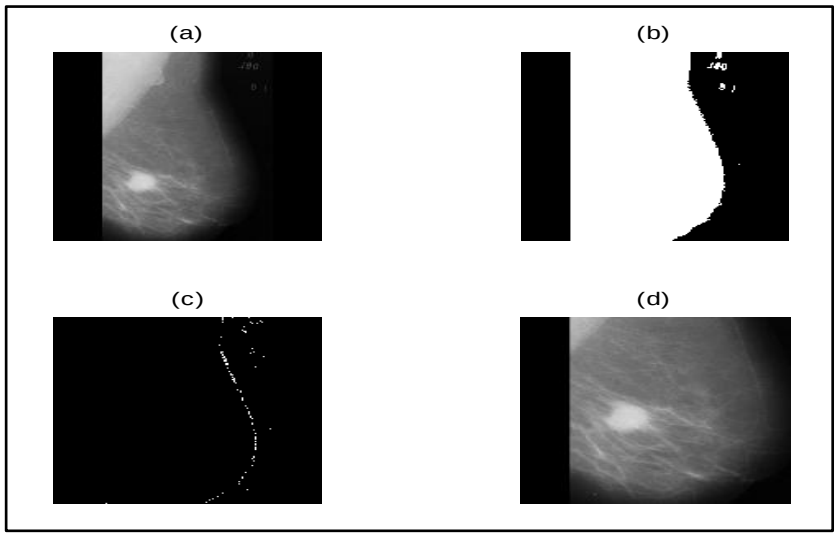

Figure 3.6 Image Pre-processing and Image Processing. (a) original image, (b) BW image, (c) *Sobel* gradient of (b), (d) Cropped image based on *Hough* parameter.

The original image (Figure 3.5.2(a)) is a grayscale image. The image then convert to BW (black and white) image (Figure 3.5.2(b)) using *im2bw* function. *im2bw (I, level)* function converts grayscale image to a binary image. The output image occur after replacing all pixels in the input image with luminance greater than level with the value 1 (white) and replaces all other pixels with the value 0 (black). The level for this function is between 0 and 1 which is relative to the signal levels possible for the image's class.

Then, the edge of the image is finding by using *Sobel* gradient as shown in Table 3.1. After *Sobel* gradient apply to image in Figure 3.5.2(b), image in Figure 3.5.2(c) obtained. By using this image, the parameters of *Hough* function and *Houghpeaks* function obtain in order to crop the image automatically and produce output image (Figure 3.5.2(d)). *Hough (BW)* computes the Standard Hough Transform (SHT) of the binary image. *Hough* function is used to detect lines in the image and generate Hough transform matrix. *Houghpeaks* function will locate peaks in the Hough transform matrix and this value will be used to crop the ROI (Region of Interest) of the image.

-1	-2	-1		-1	0	1
0	0	0		-2	0	2
1	2	1		-1	0	1

Table 3.1: Sobel approximation to the derivatives

3.2.2 System 2.

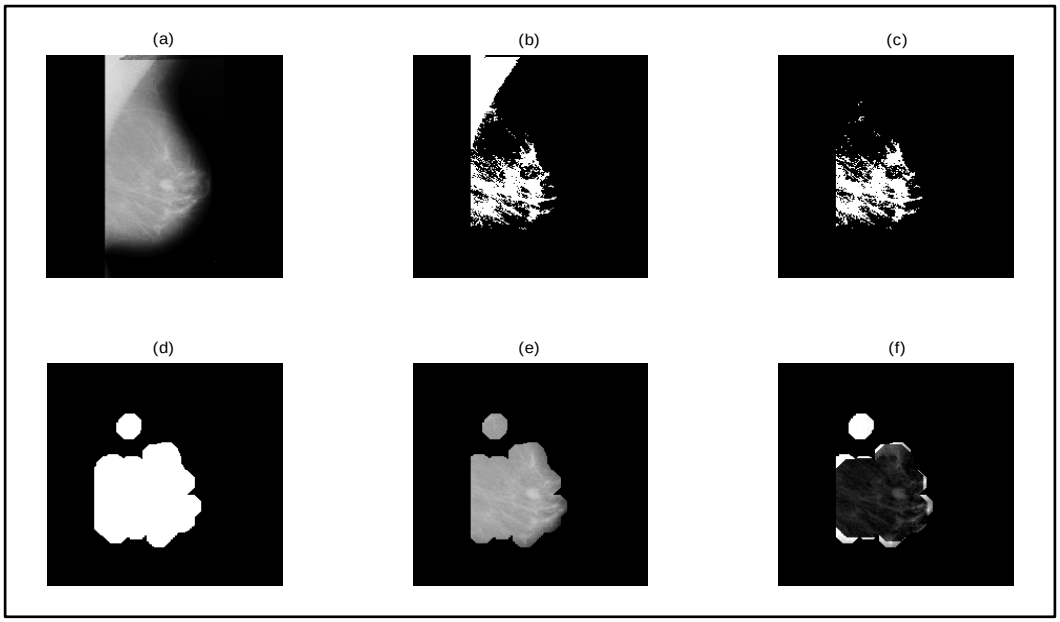

Figure 3.7 Image Pre-processing and Image Processing. (a) original image, (b) BW image, (c) *imclearborder* of image (b), (d) result after Morphological opening apply on image (c), (e) output image after multiplication between image (a) and (d), (f) sharpened image obtained by top-hat filtering and adjusting.

The image is crop to obtain new image with resolution 1001 x 1001 pixels. The images then change to BW using *im2bw* function (Figure 3.6(b)) before the noise removal function (*imclearborder*) use on the image (Figure 3.6(c)). *Imclearborder* function use to suppress structures that are lighter than their surroundings and that are connected to the image border. This step is to allow more accurate further analysis.

The image then process using morphological opening which consist of erosion followed by dilation (Figure 3.6(d)). *Imerode* function is an operation that